



NEGLIGENT AI SPEECH: SOME THOUGHTS ABOUT DUTY

*Jane Bambauer**

| | |
|--|-----|
| Introduction | 344 |
| I. Out of Scope Elements..... | 344 |
| A. Breach..... | 345 |
| B. Factual Causation and Identification of Co-Defendants..... | 346 |
| C. Proximate Causation..... | 346 |
| D. Free Speech and Section 230 Defenses..... | 347 |
| II. Duty for AI Misrepresentations Resulting in Physical Harm..... | 348 |
| A. AI Provides Information to a User that Causes Physical Harm to That User | 349 |
| 1. Model 1: “Stranger” relationship between AI producer and user | 349 |
| 2. Model 2: Special relationship between AI producer and user.. | 354 |
| 3. Model 3: AI Programs as the mischievous children or roaming animals of AI producers | 356 |
| B. AI Provides Information to a User who Causes Physical Harm to a Third Party as a Result..... | 358 |
| 1. Model 1: “Stranger” relationship between AI and user / Model 2: Special relationship between AI and user | 358 |
| C. AI Provides (or Would Have Provided) Accurate Information that Could Have Averted Physical Harm | 360 |
| Conclusion | 362 |

* Brechner Eminent Scholar, University of Florida. I am indebted to Eugene Volokh, Derek Bambauer, Alan Rozenshtein, and the other participants of this symposium for excellent and timely feedback.

INTRODUCTION

Careless speech has always existed on a very large scale. When people talk, they often give bad advice or wrong information. The scale was made more visible by the public Internet as the musings and conversations of billions of participants became accessible and searchable to all. This dynamic produced a set of tort and free speech principles that we have debated and adjusted to over the last three decades. AI speech systems bring a new dynamic. Unlike the disaggregated production of misinformation in the Internet era, much of the production will be centralized and supplied by a small number of deep pocket, attractive defendants (namely, OpenAI, Microsoft, and other producers of sophisticated conversational AI programs). When should these companies be held liable for negligent speech produced by their programs? And how should the existence of these programs affect liability between other individuals?

This essay begins to work out the options that courts or legislatures will have. I will explore a few hypotheticals that are likely to arise frequently, and then plot out the analogies that courts may make to existing liability rules.¹ The essay focuses on duty—that is, whether under traditional tort principles (which have historically accommodated and absorbed First Amendment principles)—courts should even *entertain* a case. Where there is no duty, a claim will fail early even if the plaintiff would be able to prove a lack of reasonable care, factual and legal causation, and damages.

In the end, I conclude that existing duty rules, if not modified for the AI context, could wind up missing the mark for optimal deterrence. They can be too broad, too narrow, or both at the same time, depending on how courts decide to draw their analogies.

I. OUT OF SCOPE ELEMENTS

Before launching an analysis of duty, I would like to note a few ways in which generative AI tools such as ChatGPT will raise novel issues for the other negligence elements. I address these at the outset because Duty, as the element that serves as a gatekeeper for a wide range of public policy considerations, can be affected by complications that are expected to routinely arise in the other elements. But after this

¹ I am taking off the table the possibility that the AI program itself is a “person” or entity that can bear responsibility. Nadia Banteka, *Artificially Intelligent Persons*, 58 HOUS. L. REV. (2021).

brief overview of the other elements, I will then set them aside for the remainder of the paper.²

A. *Breach*

The hypotheticals I will present in the next part assume that an AI tool will provide a recommendation or information that may affect the health or safety of the user or a third person. When might this recommendation fail to meet standards of reasonableness?

On one hand, OpenAI and other AI companies should be able to generate accurate information or useful advice *at least* as well as a person of ordinary aptitude, flawed as we mortals are in our ability to read and synthesize gobs of information. Thus, when AI programs give advice or information, one might think they should be able to do so in a manner that is at least as accurate or appropriate as a person of ordinary prudence and ability.

However, in reality there are reasons and abundant examples to illustrate that AI-generated answers will often be not just wrong but *differently* wrong, even if they are wrong less often than humans. Deep Blue's Jeopardy! experience showed us long ago that AI can hit the hard questions and miss the easy ones. ChatGPT is racking up examples today. The last six months have produced a litany of examples where ChatGPT combines a brazen level of overconfidence with a willingness to make stuff up that would make a psychopath blush. These results are not surprising given the way machine learning processes work. For instance, while a person reading up about mushrooms would be able to easily recognize that their sources make no claims at all about the safety of eating a particular mushroom while noting that others are non-toxic, an AI that uses certain large language model processing might associate the name of the toxic mushroom with the non-toxic notation of other mushrooms that happens to appear nearby.

In time, AI programs will probably be able to provide better context and confidence levels for their advice. At that point, in any given instance, it might seem fair

² I will assume that readers are familiar enough with generative AI large language models and how they work, at a basic level. I recommend other essays in this symposium for an overview. See Nina Brown, *Bots Behaving Badly: A Products Liability Approach to Chatbot-Generated Defamation*, 3 J. FREE SPEECH L. 389 (2023); Derek Bambauer & Mihai Surdeanu, *Authorbots*, 3 J. FREE SPEECH L. 375 (2023); Peter Henderson, Tatsunori Hashimoto & Mark Lemley, *Where's the Liability in Harmful AI Speech?*, 3 J. FREE SPEECH L. 589 (2023).

to hold an AI producer to a reasonableness standard (perhaps a “reasonable AI developer” standard) where errors should be expected to be less frequent than, though qualitatively different from, from the errors made by a reasonable person. As Nina Brown suggests, the standard could follow the logic of the rules in design defect claims, where plaintiffs have to prove that “there is another (even hypothetical) alternative design that would be safer than the original, but as economically feasible and practical.”³ But it will be unclear what the standard should be this early in the AI era, particularly in scenarios where generative AI produces fewer harmful errors than humans do.

B. Factual Causation and Identification of Co-Defendants

In cases where AI output causes a user to take some action that causes harm to a third party, the injured party (the plaintiff) may decide to sue the developer of the AI. In that case, the AI producer will have data that will be useful for putative plaintiffs to find the identity of a user who harmed the plaintiff, or to prove that the user actually received output from the company’s AI service that could have aided or encouraged the user to harm the plaintiff.

C. Proximate Causation

The AI companies’ access to user data will also enable the companies to know more about how users may interpret or make use of the information provided by the AI. If OpenAI has enough information to be able to infer that a user is a child, or that the user has a history of violence, or some other quality that correlates with poor judgment in the user’s reaction, that capability might alter the analysis of foreseeability of harm (both to the user and to third parties). The implication is profound: it would imply that AI companies will have proximately caused an injury if they should have recognized, based on past interactions, that the user of their service is an extremist who appears to be preparing to commit a dangerous or violent act. As long as a reasonable person would have foreseen that their conduct increased the chance of harm to another, the proximate cause element is usually met, leaving elements like “breach” and “duty” to do the work of sorting out meritorious cases where the defendant is sufficiently blame-worthy.

The proximate cause element has significant overlap with duty because foreseeability is a key factor for establishing duty as well. For example, the Restatement

³ Brown, *supra* note 2, at 68.

(Third) of Torts establishes a legal duty where a person’s conduct foreseeably increases the risk to others (as, for example, driving on a public street foreseeably increases the marginal risk to others even when it is done carefully). But to the extent duty is a hard concept for cases involving AI speech and advice, it’s difficult not because it’s difficult to foresee the risk of harm in the particular case but because it isn’t clear whether foreseeability should be *enough* to assign legal or moral responsibility for indirect downstream harm. As this essay will explain, courts will have to decide whether advice should count as an affirmative act (“conduct”) that can support legal responsibility, and whether there are any other reasons to decline to recognize a legal duty.

D. Free Speech and Section 230 Defenses

Finally, I will not independently address the impact of the First Amendment or Section 230 of the Communications Decency Act in this essay, but it is worth explaining that I don’t see either of these sources of law as offering a permanent and wholesale immunity to the producers or users of AI speech systems. Matt Perault⁴ has explained why Section 230 would not protect OpenAI. Derek Bambauer’s and Mihai Surdeanu’s essay present arguments going the other way,⁵ but they recognize significant uncertainty. In any case, lawmakers and courts are increasingly open to reducing the scope of Section 230.⁶

As for the application of the First Amendment, I have little doubt that the output of AI speech programs will be covered by free speech protections for many of the same reasons that Helen Norton, Toni Massaro, and Margot Kaminski have given.⁷ Indeed, I would go *further* than their work, as it seems clear to me that a regulation of AI speech output would aim to interfere with users-as-thinkers, and should prompt a good deal of constitutional skepticism for any regulation that is

⁴ Matt Perault, *Section 230 Won’t Protect ChatGPT*, 3 J. FREE SPEECH L. 363 (2023).

⁵ Bambauer & Surdeanu, *supra* note 2.

⁶ *Gonzalez v. Google LLC*, 2 F.4th 871 (9th Cir. 2021) (remanded by Supreme Court in light of *Twitter, Inc. v. Taamneh*, 143 S. Ct. 1206 (2023)); Allow States and Victims to Fight Online Sex Trafficking Act of 2017 (“FOSTA”), H.R. 1865 (2017).

⁷ Toni M. Massaro & Helen Norton, *Siri-ously? Free Speech Rights and Artificial Intelligence*, 110 NW. U. L. REV. 1169 (2016); Toni M. Massaro, Helen Norton & Margot E. Kaminski, *SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481 (2017).

promulgated for the protection of listeners.⁸ Nevertheless, even human speakers can sometimes be subject to liability for speech that negligently causes harm, so the First Amendment does not create anything like an absolute immunity to regulatory intervention.

This essay does not analyze free speech defenses that could clip or override state tort liability that would otherwise apply. However, courts frequently incorporate free speech values when they develop and apply tort law theories to cases involving pure expression. Thus, for example, when state courts required defamation plaintiffs to prove falsity prior to *Hepps*,⁹ they did so in part because sound principles of tort law were sensitive to the special case of speech harms.

This is true in negligence causes of action, too: what it means to “breach” the duty of reasonable care (that is, to engage in unreasonable conduct) should take the special functions of speech into account. Precautions that reduce or eliminate the chance of erroneous speech may be difficult to design given the wide range of interpretation that can apply to a particular message. And overcorrection can be costly: a liability rule may deplete the pool of available communications that are not particularly valuable to the AI producer but would be very valuable to those who want to access the material. This explains in a nutshell why courts, even under common law tort principles, imposed heightened mental state requirements for plaintiffs who sue distributors of defamatory material: distribution of speech is presumed to be “reasonable” unless the distributor is on notice of defamatory material.¹⁰

But the negligence element that most directly incorporates free speech values is duty, and this was true long before the First Amendment became the formidable source of restraint on tort law that it is today.¹¹

II. DUTY FOR AI MISREPRESENTATIONS RESULTING IN PHYSICAL HARM

The analysis of duty will be easier, or at least better organized, if we consider hypotheticals in three clusters: (A) cases where the AI gives misinformation causing

⁸ Seana Shiffrin, *A Thinker-Based Approach to Freedom of Speech*, 27 CONST. COMMENTARY 283 (2011).

⁹ *Philadelphia Newspapers v. Hepps*, 475 U.S. 767 (1986); *McCuddin v. Dickinson*, 230 Iowa 1141, 1142 (1941); *Castle v. Houston*, 19 Kan. 417, 428 (1877).

¹⁰ RESTATEMENT (SECOND) OF TORTS § 581 (1977).

¹¹ See discussion of *Yania* *infra* note 16 and accompanying text.

harm to the user; (B) cases where the AI gives misinformation causing harm to a third party (via the user's conduct); and (C) cases where the user *does not* use AI when it would have averted physical harm by providing accurate information.

A. *AI Provides Information to a User that Causes Physical Harm to That User*

Consider the following hypotheticals:

- A. AI tells an adult that it is safe to eat a mushroom that is actually toxic.
- B. AI tells a child that it is safe to eat a mushroom that is actually toxic.
- C. AI tells an adult how to alter a drug therapy to address or avoid side effects, resulting in harm from the primary disease.

1. Model 1: “Stranger” relationship between AI producer and user

To start, everyone usually has a general duty to engage in reasonable care when their conduct creates a risk of physical harm to others.¹² But would-be plaintiffs often have a harder time maintaining a negligence claim based on pure expression (books, conversations with other people, etc.) because speech does not impose a physical force or intervention.¹³ A speaker only causes physical harm if there is another person—a listener—who credits the advice or encouragement of the speech and decides to engage in physical conduct. Speaker liability is always a form of indirect liability because the plaintiff must convince the court that a speaker should be held legally responsible for the acts in the physical world that the plaintiff himself (or a third party) decided to take.

To be clear, there are circumstances where expression alone can provide the basis for liability, as when a driver causes an accident by signaling to a car behind it to go around using the lane for oncoming traffic even though he or she can see another car approaching at a fast pace.¹⁴ But these involve small scale (often one-on-one) communications where the would-be defendant has an opportunity to understand the plaintiff's specific context in real time. By contrast, communications

¹² RESTATEMENT (THIRD) OF TORTS: PHYS. & EMOT. HARM § 7 (2010).

¹³ *Holt v. Kolker*, 189 Md. 636, 640 (1948).

¹⁴ RESTATEMENT (SECOND) OF TORTS § 311 (1977); *Shirley Cloak & Dress Co. v. Arnold*, 92 Ga. App. 885, 892 (1955) (“While the defendant’s driver was under no obligation to give the plaintiff any signal at all, when he undertook to do so a duty devolved upon him to exercise ordinary care to see that the way was clear ahead for the plaintiff’s car to pass safely, and whether he did so under the circumstances is a question for the jury’s determination.”); *Miller v. Watkins*, 355 S.W.2d 1, 4 (Mo. 1962).

that are made on a large scale (such as through mass media publications) are categorically carved out of the usual negligence rules, possibly to avoid constitutional conflict with First Amendment law. In the most well-known case along these lines, *Winter v. P.G. Putnam & Sons*, the Ninth Circuit held that the publisher of a book that misidentified a poisonous mushroom as edible was found to be fully protected from products liability claims.¹⁵

What can explain the difference between the cases involving books and those involving signals to other drivers? It could be that any statement made in a form of mass communication is presumed to be speech on a matter of public concern, rather than private concern, and therefore acquires greater constitutional protection.¹⁶ But I am not convinced courts would think an obscure reference to the toxicity of a mushroom in an encyclopedia is automatically commenting on a topic of greater concern than a one-on-one conversation.¹⁷

Another explanation that leads to the same result is that courts have used no-duty rules to tackle and preempt headaches that will arise during the analysis of the other elements. Consider breach: Is it realistic to think that public law can apply a reasonable care standard to the publishers or even authors of nonfiction works without risking a serious chilling effect? The reason we do not see a lot of negligence

¹⁵ 938 F.2d 1033, 1034–36 (9th Cir. 1991). *See also* *Jones v. J.B. Lippincott Co.*, 694 F. Supp. 1216, 1216–18 (D. Md. 1988); *McMillan v. Togus Reg'l Office, Dep't of Veterans Affairs*, 120 F. App'x 849 (2d Cir. 2005) (incorrect statements about Agent Orange in National Academy of Sciences publication); *Jones v. J.B. Lippincott Co.*, 694 F. Supp. 1216 (D. Md. 1988) (poor advice about the treatment of constipation in nursing textbook); *Lewin v. McCreight*, 655 F. Supp. 282 (E.D. Mich. 1987) (published bad instructions for mixing mordant, causing an explosion); *Cardozo v. True*, 342 So. 2d 1053 (Fla. Ct. App. 1977) (poisonous ingredients listed in cook book recipe); *Alm v. Van Nostrand Reinhold Co.*, 480 N.E.2d 1263 (Ill. App. Ct. 1985) (poor instructions in how-to book about tool-making that caused injuries); *Gorran v. Atkins Nutritionals, Inc.*, 464 F. Supp. 2d 315, 326–28 (S.D.N.Y. 2006), *aff'd*, 279 F. App'x 40 (2d Cir. 2008).

¹⁶ *Waller v. Osbourne*, 763 F. Supp. 1144, 1151–52 (M.D. Ga. 1991) (“[P]lacing tort liability on some forms of protected speech would require the hopelessly complicated endeavor of differentiating between different categories of protected speech raising the possibility that the worthiness of speech might be judged by majoritarian notions of political and social propriety and morality.”) (citations omitted).

¹⁷ Indeed, there is reason to think that the “public versus private concern” issue is orthogonal to the one-to-one versus public discourse issue since the Supreme Court has recognized that one-to-one conversations can be on matters of public concern. *See Givhan v. Western Line Consol. School Dist.*, 439 U.S. 410, 414 (1979); *Rankin v. McPherson*, 483 U.S. 378, 386 (1987).

litigation when somebody is “wrong on the Internet” is because courts are likely to treat generally available speech as a “buyer beware” proposition unless the speaker gives some sort of warranty.¹⁸

The element of proximate causation would also cause headaches if negligence lawsuits were permitted to be brought against mass publishers. A plaintiff who is harmed after heeding the advice of the defendant is imposing a form of indirect liability. There is, by necessity, a decision-maker standing between the defendant’s speech and plaintiff’s harm—the plaintiff herself. Thus, when one adult recommends to another to jump off a cliff, that advice between two individuals who are not legally responsible for the health or safety of each other has been treated as not sufficient to support a duty.¹⁹ Instead, the listener is presumed to have the autonomy and responsibility to decide for themselves whether to accept, discount, or reject a recommendation.²⁰

Another reason to allow liability for some forms of speech (like hand signals while driving) while generally prohibiting liability for mass communications defendants goes back to fundamental theories of responsibility.²¹ The Supreme Court recently ruled that Internet platform services like Twitter and Google could not be

¹⁸ Tyler Cowen recently made this same point:

It is impractical to demand that all published information be the right combination of true and harmless. And what is the output of an LLM but a new and more powerful kind of book or map? (Or how about a more mischievous question: What if the LLM query requested that the answer be printed in the form of a book?)

Tyler Cowen, *Who Should Be Held Liable for AI’s Harms?*, WASH. POST (Apr. 21, 2023).

¹⁹ *Yania v. Bigan*, 155 A.2d 343, 345–46 (Pa. 1959).

²⁰ Note that there are times when courts suggest that there is a general duty to avoid negligent misrepresentation that foreseeably cause physical harm. *See, e.g.*, *Onita Pacific Corp. v. Bronson Trustees*, 843 P.2d 890 (Ore. 1992). However, these statements are made in dicta in the course of rejecting claims where negligent misrepresentation has caused only economic harms.

²¹ My preferred understanding of negligence duty in the speech context is that speech is usually not treated as a form of “misfeasance,” and is instead a form of nonfeasance. Although we have a general duty to take reasonable precautions any time our *act* creates a risk of harm, the act/omission line may have parallels to the speech/conduct line drawn in First Amendment cases. Although pure speech can foreseeably change the world, it does so exclusively through mental processes, when a listener changes their intent or comes to learn or believe something that they did not before. A book or computer may have mass and movement, but when speech causes foreseeable harm, it is because of how it affects the mental world, not the material one.

held responsible under federal statutory law for aiding and abetting terrorist organizations even though the companies knew that their services were used by members of those organizations to recruit new members who would attack and kill innocent people.²²

Importantly for our purposes, the Court reached its holding using common law tort principles related to duty.²³ The Court unanimously decided Twitter and other large Internet platforms were not consciously, voluntarily, or culpably participating in acts of terrorism because the companies did not provide ISIS users with any special functionality or promotion.²⁴ Other than the platforms' attempts to *disfavor* terrorist content through removals and suspensions, the firms treated ISIS users the same as any other user of the platform. Twitter and Google did not do anything, after establishing the functionality of the platform and recommendation algorithms, that would have associated the companies with participation or desire to bring about the specific acts of terrorism.²⁵

Open AI and the producers of other large language models are in a similar position. To be sure, these companies use software to *create* messages rather than merely hosting them. This is the reason they may not fall within the immunity of Section 230. Nevertheless, the crafting of outgoing messages is done on a large scale and based on automated procedures set in advance. This is very compatible with the reasoning in *Taamneh*. “[T]he only affirmative ‘conduct’ [is] creating their platforms and setting up their algorithms to display content relevant to user inputs and user history. . . . Once the platform and sorting-tool algorithms were up and running, defendants at most allegedly stood back and watched; they are not alleged to have taken any further action[.]”²⁶ Because the output of ChatGPT occurs without conscious oversight or awareness of individuals at the company, courts are likely to see Open AI’s role in these cases as analogous to the services of a platform like Google.²⁷

²² Twitter, Inc. v. Taamneh, 143 S. Ct. 1206 (2023).

²³ *Id.* at 1220–21.

²⁴ *Id.* at 1226–28.

²⁵ *Id.*

²⁶ *Id.* at 1226–27.

²⁷ *Id.* at 1227 (“Second, because of the distance between defendants’ acts (or failures to act) and the Reina attack, plaintiffs would need some other very good reason to think that defendants were

Thus, notwithstanding the Second Restatement of Torts' broad articulation of liability based on negligent misrepresentation,²⁸ my understanding of the caselaw is that duty in negligent speech cases is in practice drawn narrowly. Except in unusual circumstances involving speech directed to children or individuals known to have a cognitive disability, speech with strangers only supports a duty when it involves situations like signaling drivers where the defendant consciously communicates in response to a specific context, perhaps in real time, and is highly likely to induce reliance. Otherwise, tort law traditionally does not support negligence liability for bad advice outside special relationships.²⁹

So, if courts regard the producer of an AI as a "stranger" negligence case, the claims should fail for the same reason that people generally can't sue others who

consciously trying to help or otherwise "participate in" the Reina attack."), 1226 (distinguishing from a hypothetical case where the Internet company consciously selected and promoted ISIS content based on its message).

²⁸ The Restatement of Torts acknowledges that the application of liability is actually more narrow in practice:

The rule stated in this Section finds particular application where it is a *part of the actor's business or profession* to give information upon which the safety of the recipient or a third person depends. . . . The rule is not, however, limited to information given in a business or professional capacity, or to those engaged in a business or profession. It extends to any person who, *in the course of an activity which is in furtherance of his own interests*, undertakes to give information to another, and knows or should realize that the safety of the person of others may depend upon the accuracy of the information.

RESTATEMENT (SECOND) OF TORTS § 311 cmt. b (1977) (emphasis added to show that foreseeability alone is not sufficient). Comment (c) explains that liability can apply even when a negligent misrepresentation is offered gratuitously and for no self-interested purpose of the defendant, but only if the plaintiff has reason to believe the defendant has special knowledge.

²⁹ Note that the line between "stranger" cases and "special relationship" cases is blurry. Even classic cases like *Tarasoff* that require professionals to affirmatively act for the protection of strangers based on their connection to patients or clients shows that there is some flexibility and potential expansion in the category of "special relationships." *Tarasoff v. Regents of the University of California*, 17 Cal. 3d 425 (Cal. 1976). As a result, doctors and other professionals will sometimes be liable for false information even if the person harmed is somebody other than their patient or beneficiary. *See Doe v. Cochran*, 332 Conn. 325 (2019); *M.B. v. Schuylkill County*, 375 F. Supp. 3d 574 (E.D. Pa. 2019) (misrepresentation by foster child placement agency). However, these cases involve plaintiffs who are closely tied to the third party who is in a special relationship with the defendant, and as a result are in at least a quasi-special relationship with the defendant as well.

have posted false health claims on the Internet. Plaintiffs should not be able to succeed unless they can show the AI company knew or should have known that its program was communicating with a child.³⁰

To my knowledge, courts have not had to work out when that might be with respect to tort liability. An analogy to COPPA enforcement would suggest it is not enough for the AI service to be publicly available. In order for a child-plaintiff to bring a claim, they would have to show the AI service had targeted child audiences through marketing or other means.³¹ However, one open question, even if the stranger model is adopted, is whether an AI company should have to use data available to it to infer whether the user is underage or cognitively deficient and, if so, to limit its services in some manner to avoid either duty or breach. In other words, courts might interpret the fact that AI firms collect data about their end users as sufficient in some cases to create constructive knowledge that the user of the service is a child.

2. Model 2: Special relationship between AI producer and user

Unlike “stranger” cases, fiduciaries and others who are in a “special relationship” with the plaintiff owe a duty to affirmatively act on their behalf. This includes a duty to provide relevant and useful advice and to avoid negligent advice that creates risk. The usual questions about whether a putative defendant has “acted” or not are irrelevant. Instead, the inquiry would shift to whether the defendant is in a special relationship with the plaintiff.

As with the “stranger” case, proving that an AI service provider is in a special relationship with the plaintiff should be an uphill battle in most circumstances. Even if descriptively, people routinely rely on an AI service for advice related to their health, this is not sufficient on its own to establish a special relationship. Otherwise, Google would be in a special relationship with us all.³² Thus, it is instructive,

³⁰ *Yania v. Bigan*, 397 Pa. 316, 320 (Pa. 1959) (“Had Yania been a child of tender years or a person mentally deficient then it is conceivable that taunting and enticement could constitute actionable negligence if it resulted in harm.”)

³¹ Children’s Online Privacy Protection Act of 1998, 15 U.S.C. §6501 *et seq.*; 16 C.F.R. § 312.2 (COPPA regulation defining websites targeted at children).

³² For arguments that this would or would not be a desirable evolution in the law, see Jack Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016), and Jane Bambauer, *The Relationships Between Speech and Conduct*, 49 U.C. DAVIS L. REV. 1941 (2016), respectively.

when contemplating AI liability of various sorts, whether the law should apply any differently to the results of an AI prompt as it does to the results of a Google search that points to bad content on the Internet.

However, courts may seize on the distinction that (1) Google's service points people to information that typically form the inputs to a user's decision-making process while (2) in the near future, users may rely so completely on the synthesis that an AI program provides that they essentially outsource the decision-making process. If AI users rely on AI for the *output* of a decision-making process, bypassing the part where they weigh and interpret the information for themselves, courts may find this pattern of behavior calls for the recognition of a special relationship.

Courts should be wary of accepting this argument because legal rules and human behavior are endogenous. If courts establish that general purpose AI producers are in the equivalent of a special relationship with users because they provide advice that is followed without independent and autonomous choice of the user, that rule will *induce* dependence and reliance. This is appropriate in some contexts, but not in every context where human health and safety are involved.

That said, if a producer makes special-purpose AI oriented toward providing guidance in specialized domains like law or medicine, it may very well meet the requirements of a special relationship under the common law. For better or worse, courts will probably look at factors such as whether the service provider charges prices consistent with professional services, whether the service is marketed with reference to these sorts of domains of professional expertise, and whether users trust the service with consequential decisions. Thus, hypothetical (C), where an AI chatbot makes recommendations for altering a drug therapy, is most likely to fall within an area of traditional tort duty if the producer provides a specific-purpose, rather than general-purpose, AI service. In that case, a producer may have to comply both with negligence law that traditionally applies to doctors *and* with FDA regulation that traditionally applies to devices.³³

The fact that this is so—that the same service might be categorized as both an (artificial) doctor and as a device—demonstrates that AI may blur doctrinal lines that were once clean, such as the line separating products liability from professional

³³ U.S. FOOD & DRUG ADMIN., CLINICAL DECISION SUPPORT SOFTWARE: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2022), <https://perma.cc/G6F6-39M5>.

malpractice. Conceptualizing it as the former—as a product—would allow defendants to tap into the logic of cases like *Winter* that categorically exempt speech products from products liability, and where traditional duty rules have taken a “buyer beware” stance for listeners. But if AI is analogized to a professional advisor, then AI companies will be virtually certain to have a professional duty of care.

3. Model 3: AI Programs as the mischievous children or roaming animals of AI producers

Another useful, if strained, analogy for generative AI is that of wandering precocious children. By this analogy, OpenAI is the absent parent of a young child who can't be expected to adhere to an adult standard of care. Of course, adults (and even other children) who engage with young children are on notice that the young child's actions and communications are ill-conceived and unpredictable. This is less so with generative AI output. Open AI encourages users to take the Chat GPT output seriously, and even if it didn't, the fact that so much of the output is correct and sound makes the AI hallucinations all the more insidious and harder to guard against.³⁴ Thus, it acts more like a child who is a prodigy who knows more than the adults who ask him questions, but every once in a while slips in a complete fabrication.

Nevertheless, the comparison is useful because it could help courts divide problems into those where the producers of AI can be sued under theories of “negligent supervision.”³⁵ Parents are expected to anticipate where and when their child might be doing dangerous things and to supervise and intervene, if necessary, in those contexts. Thus, against a general backdrop of no duty, a parent will be expected to use care when providing their child with a motor vehicle or gun.³⁶ For younger children, parents are expected to watch somewhat more vigilantly when their child is

³⁴ OPEN AI, GPT-4 TECHNICAL REPORT 46 (2023), <https://arxiv.org/pdf/2303.08774.pdf> (“This tendency [to hallucinate] can be particularly harmful as models become increasingly convincing and believable, leading to overreliance on them by users. . . . Counterintuitively, hallucinations can become more dangerous as models become more truthful, as users build trust in the model when it provides truthful information in areas where they have some familiarity.”)

³⁵ As a bonus, the adoption of negligent supervision frameworks will permit many corny puns on “supervised learning”—an important part of the machine learning process during the development of AI.

³⁶ *Moore v. Crumpton*, 306 N.C. 618, 623 (1982) (“The correct rule is that the parent of an unemancipated child may be held liable in damages for failing to exercise reasonable control over

interacting with other vulnerable children in pools or on playgrounds, but there is also an understanding that accidents will happen. A successful negligent supervision case requires something more than a theory that the supervisor could have exerted more control over the children.³⁷ Courts that conceive of harmful AI output cases as a category of negligent supervision by the producer of the AI program would focus less on the specific facts leading to injury and more on the category of output and the feedback mechanisms that the AI producer has put in place to detect recurring problems.

Old common law rules related to dogs and cattle may lead to a similar result. When animals roam onto others' property and cause damage or physical harm, courts don't look to the mental state of the animal, of course, even though it is sentient to some degree. But they also do not completely ascribe the actions of the animal to its owner. Especially in the 19th century, when much of the economy in less populous states revolved around farming, courts that adopted the open range / fence out rules understood that the owners of cattle could not fully control something that had a mind of its own. The fact that jurisdictions adopted different duty rules depending on the population—some using no duty (open range), some using negligence, and some adopting strict liability—demonstrates that higher order cost-benefit analyses often play out in the *duty* element rather than in breach. Reduced-duty rules for dogs like the proverbial “one bite rule” can be understood the same way.³⁸

The point is that courts have long had to deal with autonomous things that are welcome parts of society but can cause harms that they cannot themselves be held

the child's behavior if the parent had the ability and the opportunity to control the child and knew or should have known of the necessity for exercising such control.”).

³⁷ *Norman v. Turkey Run Community School Corp.*, 274 Ind. 310, 316 (1980) (“No teacher can observe every student at every instant on a playground. To look at one is to look away from another. Even if the evidence showed that one or both teachers were looking in another direction, it would not give rise to an inference of negligence on the part of either or both of them.”).

³⁸ *Modern Status of Rule of Absolute or Strict Liability for Dogbite*, 51 A.L.R.4th 446 (1987) (“The cliché that a dog is allowed one bite has grown increasingly inaccurate as many jurisdictions have adopted various forms of absolute or strict liability as to dog owners, through either statutory enactment or interpretation of common law. However, despite this tendency to label a dog owner's liability as absolute or strict, the actual or constructive knowledge of the owner of the animal's viciousness or dangerous propensities frequently remains an essential element.”).

accountable for. The nearest-responsible adult humans are not usually held vicariously responsible or strictly liable, but neither are they immune from liability. Instead, duty rules are conscribed to special sets of facts. And the analogy is also useful in answering the question: to the extent AI is like children or cows, who is the equivalent of the AI's supervisor? Probably usually the producer, who sets fences and monitors improvements, but perhaps there will be some instances where the user of the AI has taken on a supervising role, i.e. by using a generative AI to create new programs or agents at the user's direction.

B. AI Provides Information to a User who Causes Physical Harm to a Third Party as a Result

Next, consider these scenarios:

- A. AI tells a user that a certain toxic mushroom is edible, and he shares a mushroom with his friend or child.
- B. AI correctly tells a user that a certain mushroom is toxic, and the user maliciously poisons a neighbor.
- C. AI provides incorrect or incomplete information to a doctor about a treatment plan for a patient with a unique medical history, and the doctor follows the AI's advice to the patient's detriment.

1. Model 1: "Stranger" relationship between AI and user / Model 2: Special relationship between AI and user

The "no duty" rule that typically applies in cases where a listener harms himself does not necessarily apply where the listener foreseeably harms a third person. If a person's speech encourages another to take an action that will put another person in peril, the victim might be able to seek recovery from the person who directly caused the injury (the listener) *and* from another who induced the listener to act through speech. In *Weirum*, for example, the relatives of a man injured in a car accident caused by speeding teenagers successfully sued the radio station that encouraged the teenagers (and other listeners) to find a DJ's car as quickly as possible.³⁹

³⁹ *Weirum v. RKO General, Inc.*, 15 Cal. 3d 40, 47 (1975). *See also* *Stricklin v. Stefani*, 358 F. Supp. 3d 516, 529 (W.D.N.C. 2018) (finding that Gwen Stefani owed a duty of care when she told the audience to move toward the stage).

The *Weirum* court stressed that its outcome was dependent on the unusual facts, such as the design of a game that would be played in real space, and the provision of a valuable prize for the winners. The Georgia Supreme Court recently reached a similar result in a case brought against Snapchat based in the design of its “speed filter” because the company allegedly “knew that other drivers were using the Speed Filter while speeding at 100 miles per hour or more as part of ‘a game,’ [and] purposefully designed its products to encourage such behavior[.]”⁴⁰

But these cases tend to focus on purposeful encouragement and non-expressive elements like prizes. They are the exceptions that proves the rule: courts usually dismiss cases attempting to assign negligence liability based on speech alone.⁴¹ Cases brought against producers of books and rap songs that encourage violence are losers for the plaintiffs even when it is foreseeable, in a probabilistic sense, that a listener might engage in copycat crimes.⁴² In *Rice v. Paladin Enterprises*, a rare case of publisher liability, the Fourth Circuit permitted a case against the publisher of a hit man “how to” instruction manual to proceed because the publisher “not only knew that its instructions might be used by murderers, but [] actually intended to provide assistance to murderers and would-be murderers which would be used by them[.]”⁴³ Scenario B, therefore, seems to be completely foreclosed unless the

⁴⁰ *Maynard v. Snapchat*, 313 Ga. 533, 540 (2022). Courts have also permitted tort suits brought by patients who suffered harm from severe side effects against pharmaceutical manufacturers that failed to inform the plaintiff’s doctors about the side effects. *McEwen v. Ortho Pharmaceutical Corp.*, 528 P.2d 522, 528–29 (Or. 1974). These types of cases, however, are extensions of the special relationship style of case. Liability is contingent on the company having a special relationship and owing an affirmative duty to warn to *somebody* (the doctors) and causing foreseeable harm to others as a result of breaching their duty.

⁴¹ *Rubio v. Swiridoff*, 165 Cal. App. 3d 400, 405 (1985) (rejecting the notion that the defendant owed a duty to a third party when he got in a heated argument with his girlfriend and she recklessly drove her car into the plaintiff’s).

⁴² *Olivia N. v. National Broadcasting Co.*, 126 Cal. App. 3d 488, 495 (1981); *Herceg v. Hustler Magazine, Inc.*, 814 F.2d 1017, 1020 (5th Cir. 1987); *Yakubowicz v. Paramount Pictures Corp.*, 404 Mass. 624, 631 (Mass. 1989); *Waller v. Osbourne*, 763 F. Supp. 1144, 1151 (M.D. Ga. 1991); *Bill v. Superior Court*, 137 Cal. App. 3d 1002, 1011 (1982) (“Here, by contrast to *Weirum*, the petitioners’ activity in producing a motion picture and arranging for its distribution, is socially unobjectionable—and, in light of First Amendment considerations, must be deemed so even if it had the tendency to attract violence-prone individuals to the vicinity of theaters at which it was exhibited.”).

⁴³ *Rice v. Paladin Enterprises, Inc.*, 128 F.3d 233, 242 (4th Cir. 1997).

plaintiff can show that an AI producer *wanted* their output to facilitate law-breaking.⁴⁴

For the other hypotheticals, the outcome of cases will strongly depend on whether courts analogize to pure speech products, to strangers, or to professional advisors. The book analogy will lead to a no duty rule. The stranger analogy will require courts to assess how specifically foreseeable it may have been that a listener would take some imminent action that creates risks to others as a result of the communication. And the advisor/fiduciary analogy will establish duty in almost every case.

C. AI Provides (or Would Have Provided) Accurate Information that Could Have Averted Physical Harm

Finally, let's consider how the liability of *users* might change in light of the ready availability of AI tools.

- A. AI provides information to a doctor about a treatment plan for a patient with a unique medical history, and the doctor *does not* follow the AI's advice.
- B. The doctor does not consult an AI at all, and thus fails to learn about the optimal treatment plan.
- C. A person shares a mushroom with his friend or child without consulting a readily available AI to determine whether it is poisonous.

In all three of these hypotheticals, the user's duty towards the plaintiff is easily established. In hypos (A) and (B) there is a doctor-patient relationship, and in hypo (C) the act of serving food is sufficiently affirmative to establish the general duty of reasonable care.

In each of these, it is possible that the existence of a high performance AI resource would alter the analysis of breach such that a doctor or the mushroom sharer can be found to fall below the reasonable care standard where they may not have been in breach in the past.

⁴⁴ Twitter, Inc. v. Taamneh, 143 S. Ct. 1206, 1220 (2023) ("Importantly, the concept of 'helping' in the commission of a crime—or a tort—has never been boundless. That is because, if it were, aiding-and-abetting liability could sweep in innocent bystanders as well as those who gave only tangential assistance. . . . Other cases have emphasized the need for some 'culpable conduct' and 'some degree of knowledge that [a defendant's] actions are aiding the primary violator' before holding the defendant secondarily liable.").

This dynamic is not unusual: information and communications technologies frequently push standard of care up over time. But does this suggest that the common law is at war with itself, discouraging activity among AI developers in some cases (by establishing duties of care) while encouraging it in others (by creating de facto requirements that doctors, e.g., use AI)?

I don't think so. In fact, the potential liability that some may in the future face for *not* using AI is an important feature that can help ensure the common law does not deter AI innovations that will bring net improvements in health and human welfare.⁴⁵

Consider, for example, the doctor who does *not* consult with generative AI services while managing a complex case. Medical researchers have already found that Open AI's GPT-4 outperforms human doctors in some therapeutic decision-making tasks.⁴⁶ That means it is plausible that there may be a viable medical malpractice claim brought against a medical doctor within the next year or two that tests the theory the doctor acted unreasonably by failing to seek out information from ChatGPT. This could be a plausible claim whether the jurisdiction uses custom as the standard of care or instead uses cost-benefit methods of establishing reasonable care.

These interlocking duties—where tort law may steer individuals to adopt AI assistance faster while also imposing responsibilities on AI developers to avoid foreseeable and needless risks—make tort law a good vehicle for risk management in the context of rapid innovation. The deployment and mass adoption of AI could reduce all sorts of risks that today we call “baseline.” Tort law has the flexibility to recognize liability rules where AI developers needlessly inject risk while also imposing costs on people who needlessly avoid AI's help.

⁴⁵ For this reason, the fact that the FDA has claimed jurisdiction over all software systems that make recommendations to patients should raise concerns of over-regulation. See David A. Dorr et al., *Harnessing the Promise of Artificial Intelligence Responsibly*, JAMA (Mar. 27, 2023).

⁴⁶ Eric Topol, *The GPT-x Revolution in Medicine*, GROUND TRUTHS (Mar. 27, 2023) (reviewing a forthcoming book finding that although hallucinations are a problem, GPT-4 performs well for both digesting information and explaining diagnoses to patients). For a less sanguine account, see Dev Dash et al., *How Well Do Large Language Models Support Clinician Information Needs?*, STANFORD UNIVERSITY HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (Mar. 31, 2023) (finding that while the vast majority of tested responses were safe (in the sense of not likely to cause harm), only 41% agreed with a known answer).

CONCLUSION

At this early stage of development and critical period of maturation, courts will be under tremendous pressure from consumer advocates and the plaintiffs' bar to impose duties of care on the large companies that are developing generative AI services. At the same time, courts will also anticipate the pitfalls from assigning duties of care too expansively, even when AI speech causes physical injuries. Although generative AI is a *sui generis* phenomenon that isn't perfectly analogous to anything that has come before, courts can still make use of analogies to past scenarios that share certain qualities with the products and services of AI companies.

Courts are probably best served by starting with the presumption that AI output should be treated the same as Google search results or mass media products, as far as tort duties are concerned. When AI speech causes physical harm, it does so indirectly and through pure expression produced at a scale that cannot be comprehensively monitored in real time. However, courts should embrace analogies to duty rules that apply to the parents of mischievous children or to the owners of animals in contexts where AI producers can anticipate recurring problems and could implement guardrails without undermining the quality of the service. And analogies to doctors, lawyers, or other fiduciaries (who of course do owe duties of care) will be appropriate for special use AI programs that are specifically marketed and designed to perform an advisor function—presumably at a price point that is much greater than current ChatGPT prices.